Relationships between Investments in Science and Scientific Output: Evidence from Cross-National Panel Data

Joseph M. Staudt
The Ohio State University
staudt.8@osu.edu
May 2010

Abstract

Economists have long emphasized the important role that new ideas play in promoting economic growth. Since formal scientific discovery is an important source of new ideas, it plays a potentially foundational role in the growth process. Due to its importance, a literature on the determinants of scientific output and productivity has emerged. Most of this literature focuses on the individual researcher or institution as the unit of analysis. Very little attention has been paid to the determinants of scientific output at the country-level. Toward a better understanding of the country-level relationship between investment in science and scientific output, this paper uses panel data from the World Bank and OECD to estimate the elasticity of scientific output with respect to investment in science. Our estimates range from 0.25 to 0.71. In addition to these contemporaneous estimates, we present evidence that past investment is also related to output. We conclude that there is an economically significant positive association between investment in science and scientific output. A sensitivity analysis reveals that this conclusion is quite robust. However, we make no attempt to establish causality, and problems with the data and econometric difficulties dictate that our estimates should be interpreted with caution.

I am very grateful to Bruce Weinberg, whose encouragement and guidance made this project possible. Also, thanks to Masanori Hashimoto and Craig Boardman for serving on my thesis committee and offering useful comments. Finally, thanks to the OSU Arts and Sciences and OSU College of Social and Behavioral Sciences for financial support of my work. I am solely responsible for all errors.

I. Introduction

One of economists' primary motivations for studying science is its potential impact on economic growth (Stephan, 1996). Since the advent of endogenous growth theory in the late 1980's, economists have put more emphasis on the role that new knowledge plays in promoting growth [for instance, see Romer (1986) and Lucas (1988)]. Because formal scientific discovery is an important source of new knowledge, it plays a potentially foundational role in the growth process, and so naturally commands the attention of economists. Indeed, economists and other social scientists have put forth much effort to understand the determinants of scientific output and productivity (Boardman, forthcoming). Most of this research focuses on individual researchers or institutions as the unit of analysis. However, with the exception of Crespi and Guena (2004) and Crespi and Geuna (2005), there has been surprisingly little research conducted at the country level. Tellingly, neither Stephan (1996) nor Diamond (1996) mentions cross-country estimates of the return to scientific investments in their broad surveys of the economics of science.

This lack of country-level research is all the more puzzling in light of the fact that policymakers and members of the business community have recently been allocating larger shares of budgets to investments in science (as measured by total R&D expenditure). As figure 1 shows, growth rates in total R&D expenditure have been quite high in most developed countries. Between the years 1990 and 2000, the mean real annual growth rate for total R&D expenditure in developed countries was about 4 percent. This growth rate outpaced GDP growth causing an increased proportion of resources to be allocated to R&D. As developed countries continue to emphasize the importance of scientific investment as a foundation of national competitiveness and developing countries increasingly turn to science as a means of reaching the economic frontier, a better understanding of the relationship between investments in science and scientific output is imperative to devising sound policies concerning the most effective ways to fund scientific research.

This paper takes a step toward better understanding the relationships between investment in science and scientific output by using cross-national panel data from the World Bank and the Organization for Economic Cooperation and Development. Consistent with most of the existing literature, we find that there is a strong positive relationship between investments in science and scientific output. In particular, estimates of the elasticity of scientific output with respect to investments in science range from about 0.25 to 0.71. In addition, there is also evidence that past investment in science is related to scientific output. Estimates of the elasticity of output per researcher with respect to per researcher investment in science range from 0.30 to 0.57. Unlike total output, we find little evidence that output per researcher is affected by past investments. As will be shown, these results are robust across many model specifications and estimation techniques.

It is important to emphasize that our objective is the *description* of the relationship between scientific investment and scientific output. Due to the lack of a satisfying instrumental variable, we make no attempt to establish *causality*. However, we do discuss several possible causal mechanisms that could have produced our results. Furthermore, due to data inadequacies, and possible simultaneity or omitted variables, our estimators may be biased.

The rest of this paper is organized as follows. Section II provides an overview of the literature concerning the determinants of scientific output and productivity. Section III provides a description of the data we use to analyze the relationships between investment in science and scientific output. Section IV discusses the conceptual framework underlying our empirical strategy. Section V discusses the econometric models we use to analyze the relationship between investment in science and scientific output, as well as the techniques we use to estimate the models. Section VI presents the results of our estimations, provides a sensitivity analysis to examine the robustness of our results, discusses causality, and provides a discussion of the shortcomings of our analysis. Finally, section VII concludes.

II. Literature Review[1]

Sociologists, lead by Merton (1973), have a long history of studying the determinants of scientific output and productivity. Most of this work has been qualitative and the quantitative work that has been done has not focused on the relationship between investment in science and scientific output. Instead, the quantitative sociological literature has focused on potential correlates of productivity such as institutional type, gender, age, and satisfaction with the promotions system. For instance, see Pelz and Andrews (1966), Long and McGinnis (1981), Fox (1983), Ramsden (1994), and Teodorescu (2000).

Economists have put more emphasis on the relationship between investment in science and scientific output, using a variety of measures for each. They have also put more emphasis on quantitative methods of analysis. For example, Adams and Griliches (1996) used data on American universities during the 1980's to examine the relationship between R&D expenditures and research output, as measured by papers and citations, in eight different fields of science. At the field level, they find parity between growth rates of R&D expenditure and growth rates for papers and citations for six of the eight fields analyzed. At the university level, however, the authors find that returns to R&D expenditures are diminishing in nearly every field. Likewise, Payne and Siow (1998) used panel data on federal research funding to 71 American research universities and found that increased funding is associated with increased research output. However, they found that increased output does not necessarily translate into higher quality

[1] For a more comprehensive literature review, see Boardman (forthcoming).

output. Coupé (2003), using patents as a measure of research output, found that increased funding for academic research increases the number of university patents. Finally, using data on Nobel laureates in chemistry, medicine, and physics, along with data on highly cited publications, Weinberg (2008) suggested that investments in science were one of the primary factors that enabled the United States to displace Europe, and Germany in particular, as the world leader in scientific output during the 20th century.

Although there is a literature on the determinants of scientific output, almost all research is confined to analyses within countries. More specifically, individual researchers, scientific fields, or institutions within a country are usually the unit of analysis. There has been surprisingly little work done at the country-level. To our knowledge, only Crespi and Geuna (2004) and Crespi and Geuna (2005) have attempted to systematically analyze the relationship between investment in science and scientific output from a cross-national perspective. Thus, much work remains to be done on the determinants of scientific output and productivity in a cross-national context. This paper contributes to this literature by estimating the elasticity of scientific output with respect to investment in science using a variety of model specifications and estimation techniques.

III. Data

To estimate the relationship between scientific investment and scientific output we use data from the World Bank's World Development Indicators (WDI) data set and the Organization for Economic Cooperation and Development's (OECD) Main Science and Technology Indicators data set. Specifically, we constructed an unbalanced panel data set covering 83 geographical units between the years 1990 and 2005. Table 1 lists the countries used in the analysis.

The data set is comprised of four main components: a) two measures of scientific output, b) one measure of scientific output per researcher, c) two measures of investment in science, and d) various control variables. The two measures of scientific output are the number of scientific journal articles published by researchers within a country and the number of R&D researchers within a country. The measure of output per researcher is the number of articles divided by the number of researchers. Output per researcher can be thought of as a measure of scientific productivity.

Notice that, to some extent, the number of researchers can be thought of as an output or an input. We can justify using the number of researchers as a measure of scientific output because a large share of scientific investment is used to fund the education of new researchers. On the other hand, we can think of the number of researchers as a measure of the labor input into the scientific production process. When we consider the number of researchers as an input, we can divide the number of articles by the number of researchers to produce a measure of output (articles) per unit of input (researchers). In this sense, our measure of output per researcher can be thought of as a measure of scientific productivity.

The two measures of output and one measure of output per researcher are from the WDI data set.[2] Note that one of the major shortcomings of this analysis is that there are no controls for the *quality* of articles or researchers. This issue is addressed further in section VI.

The two measures of investment are total R&D expenditure and R&D expenditure per researcher. The R&D expenditure data is from the Main Science and Technology Indicators data set.[3] The number of researchers, used to construct the variable R&D expenditure per researcher, is from the WDI data set. R&D expenditure is measured in constant 2000 US dollars.

[2] http://data.worldbank.org/indicator
[3] http://puck.sourceoecd.org/vl=2030475/cl=12/nw=1/rpsv/ij/oecdstats/16081242/v207n1/s1/p1

The most important control variables are GDP per capita, population, expenditure on education, and military expenditure. All of these variables are from the WDI data set. Table 2 lists and describes the output, output per researcher, investment, and control variables. Table 3 provides summary statistics for each variable.

IV. Conceptual Framework

Following Griliches (1979) and later Crespi and Geuna (2004), we can specify a simple knowledge production function as follows. Let $Y = F(X, S, \varepsilon)$ be the production function relating some measure of knowledge output Y (at the country-level) to inputs X, S, and ε. We let X stand for conventional inputs such as labor and other control variables. S is a measure of the current state of scientific knowledge, which is partly determined by past and current R&D expenditures. Finally, ε stands for all unmeasured determinants of output and productivity.

For convenience, we assume that F is a Cobb-Douglas and that ε is random. Including a time trend term, we can rewrite $F(.)$ as,

$$Y_{c,t} = A_c X_{c,t}^{\alpha} S_{c,t}^{\beta} e^{\gamma t + \varepsilon_{c,t}}. \quad (1)$$

Here, c is the country index and t is the time index. α and β are output elasticities to be estimated. γ is the coefficient for the time trend. A_c is a constant at the country-level. e is the natural logarithm base. Taking the natural logarithm of both sides of equation 1, this function form can be estimated as a linear relationship using the following expression,

$$\ln(Y_{c,t}) = \ln(A_c) + \alpha \ln(X_{c,t}) + \beta \ln(S_{c,t}) + \gamma t + \varepsilon_{c,t}. \quad (2)$$

The next section discusses the specific econometric models and methods that we will use to estimate equation 2.

V. Econometric Models and Estimation Methods

This section describes the general econometric models we use to describe the relationship between scientific investment and scientific output. It also describes the strategies we will use to estimate the parameters of these models. Consider a general model,

$$output_{c,t} = \delta_0 investment_{c,t} + \sum_{j=1}^{J} \delta_j investment_{c,t-J} + \boldsymbol{\beta}'\boldsymbol{X_{c,t}} + \boldsymbol{\gamma}'\boldsymbol{T_t} + \alpha_c + \varepsilon_{c,t}. \qquad (3)$$

Here, $output_{\mathrm{c,t}}$ is a measure of total scientific output for country c in year t. This term corresponds to $Y_{c,t}$ in equation 1. As noted in section III, we consider two measures of total scientific output: a) the number of scientific journal articles published and b) the number of researchers working in R&D. Both measures are in natural logarithmic form.

A country's total scientific investment in year t is given by $investment_{c,t}$. Total scientific investment is measured by the natural logarithm of total R&D expenditure. The variables $investment_{c,t-1}, \dots, investment_{c,t-J}$ represent a country's total scientific investment for years $t-1$ to $t-J$. It is plausible that scientific investment affects scientific output with a lag. Adding J lags to our general model accounts for this reality. These investment variables correspond to $\mathrm{S}_{\mathrm{c,t}}$ in equation 1.

The variable $\boldsymbol{X_{c,t}}$ is a vector of time-varying country-specific covariates. This vector includes covariates such as GDP per capita and education expenditure. These variables correspond to $\mathrm{X}_{\mathrm{c,t}}$ in equation 1. We consider a range of specifications for this term. The variable $\boldsymbol{T_t}$ is a vector of time-effect dummy variables for the years 1991 to 2005. These dummy variables correspond to t in equation 1. Country fixed effects for country c are given by α_c, which corresponds to A_c in equation 1. The error term is given by $\varepsilon_{c,t}$. As noted, table 2 describes all of the variables used in this paper.

To estimate the elasticity of scientific output with respect to total investment in science, we first consider two special cases of model 3, and then consider the full model 3. For the first special case of model 3, we set $\delta_1 = \cdots = \delta_J = 0$ and $\boldsymbol{\beta}' = \mathbf{0}$. That is, we exclude the lagged investment variables and the time-varying country-specific covariates from the model. Using the natural logarithm of the number of articles as the measure of scientific output and the natural logarithm of total R&D expenditure as the measure of investment in science, we use the random effects (RE) estimator and the fixed effects (FE) estimator to estimate this particular specification of model 3. We then change the measure of scientific output to the natural logarithm of the number of researchers. The RE and FE estimators are again used to estimate the newly specified model 3. We obtain four estimates for the first special case.

For the second special case of model 3, we allow the time-varying country-specific covariates, represented by the vector $\boldsymbol{X_{c,t}}$ to enter the model while retaining the restriction $\delta_1 = \cdots = \delta_J = 0$. Similar to our strategy for estimating the first special case of model 3, the natural logarithm of the number of articles is used as the measure of scientific output and the natural logarithm of total R&D expenditure is used as the measure of investment in science. The RE and FE estimators are then used to estimate the model. The measure of scientific output is then changed to the natural logarithm of the number of researchers. The RE and FE estimators are again used to estimate the newly specified model 3. In total, we obtain four estimates for the second special case.

Finally, we allow both the time-varying country-specific covariates and the lagged investment variables to enter the model. This gives us the general model 3, where we set $J = 5$. Similar to the special cases, we first use the natural logarithm of the number of articles as the measure of scientific output and the natural logarithm of total R&D expenditure as the measure of investment in science. The RE and FE estimators are then used to estimate model 3. The measure of scientific output is then changed to the natural logarithm of the number of researchers, and RE and FE are again used to estimate the newly specified model 3. This in results four estimates of the general model 3. Overall, there are 12 estimates of the elasticity of total scientific output with respect to total investment in science.

Note that model 3 represents the relationship between *total* investment in science and *total* scientific output. We now consider models that describe the relationship between *per researcher* scientific investment and *per researcher* scientific output. Recall that for model 3, we considered the number of researchers to be an *output*. We now consider the number of researchers to be an *input* to the scientific production process. Thus, scientific output per researcher can be considered a measure of scientific productivity. To estimate the elasticity of output per researcher with respect to investment per researcher we change the dependent variable in model 3 to the number of articles per researcher and the measure of investment in science to R&D expenditure per researcher. Both measures are in natural logarithmic form.

Our strategy to estimate the elasticity of output per researcher with respect to investment per researcher parallels the strategy we used to estimate model 3. We consider the two special cases, and then consider the full model. In each case, we use RE and FE estimators to estimate the parameters of the model. Overall, there are 6 estimates of the elasticity of scientific output per researcher with respect to per researcher investment in science. The estimates for all models are reported in the next section.

VI. Empirical Results and Discussion

Table 4 presents the estimates for each case of model 3 when the natural logarithm of the number of articles is the measure of scientific output. Columns 1 through 3 present the random effects (RE) estimates and columns 5 through 7 present the fixed effects (FE) estimates. Overall, we see that the contemporaneous estimates of the elasticity of the number of articles published with respect to total R&D expenditure range from 0.25 to 0.71. Each of these estimates is positive, economically significant, and precisely estimated.

Notice that the lagged investment variables in columns 3 and 7 are usually statistically insignificant and often have negative coefficients. Taken literally, these results suggest that past investments in science have at best no impact on scientific output and at worst a negative impact. However, there is likely to be severe collinearity between the lags.

To further probe the relationship between past investments in science and scientific output, we averaged the five lagged investment variables, creating a single new variable. We then replaced the five individual lagged variables in model 3 with the five-year averaged lag variable, and re-estimated the new model.[4]

Columns 4 and 8 of table 4 present the estimates of this newly specified model. We see that the contemporaneous estimates of the elasticity of the number of articles with respect to total R&D expenditure are smaller than our previous estimates. Nevertheless, they are basically consistent with our previous estimates. However, it is notable that the estimates of the elasticity of the five-year averaged lag variable are positive, economically significant, and precisely estimated. Thus, this new model offers evidence that past investments in science are positively associated with current scientific output, even holding current investment constant.

Table 5 presents the estimates for each case of model 3 when the natural logarithm of the number of researchers is the measure of scientific output. Columns 1 through 3 present the RE estimates and columns 5 through 7 present the FE estimates. Overall we see that the

[4] Note that the new models created by replacing the five lagged investment variables with the average of the five lags are not special cases of models 3 and 4. Rather they are entirely new models.

contemporaneous estimates of the elasticity of the number of researchers with respect to total R&D expenditure range from 0.51 to 0.68. Again, each of these estimates is positive, economically significant, and precisely estimated.

Like our estimates in table 4, the lagged investment variables in columns 3 and 7 of table 5 are statistically insignificant and often have negative coefficients. Thus, we again replace the five individual lagged variables with the five-year averaged lag. The contemporaneous estimates of the elasticity of the number of researchers with respect to total R&D expenditure are basically consistent with our original estimates. Also, we again see that our estimates of the elasticity of the five-year averaged lag variable are positive, economically significant, and precisely estimated. Thus, we find more evidence that past investments in science are positively associated with current scientific output.

Table 6 presents the estimates for the model in which the dependent variable is output per researcher (articles per researcher) and the measure of investment in science is investment per researcher (R&D expenditure per researcher). Recall that under this specification, the number of researchers is considered an input to the scientific production process and so output per researcher can be thought of as a measure of scientific productivity.

Overall, the contemporaneous estimates of the elasticity of the number of articles per researcher with respect to per researcher R&D expenditure range from 0.30 to 0.48. These estimates are all positive, economically significant, and precisely estimated. In columns 4 and 8, we replace the five individual lagged variables with the five-year averaged lag, and re-estimated. The contemporaneous estimates of the elasticity of articles per researcher with respect to R&D expenditure per researcher are consistent with our original estimates. Though our RE estimate of the elasticity of the five-year averaged lag variable is positive, economically significant, and precisely estimated, the FE estimate is negative and statistically insignificant. Since a Hausman (1978) test offers marginal evidence that the random effects model is inappropriate, we do not have strong evidence that past investments in science are positively associated with current output per researcher.

Overall, whether we use articles or researchers as the measure of scientific output, we find a positive, economically significant, and precisely estimated relationship between output and investment in science. Moreover, we find some evidence that past investments impact current output, even holding current investments constant. In addition, we find a positive, economically significant, precisely estimated relationship between output per researcher (which can be thought of as a measure of productivity when we consider researchers an input into the scientific production process) and per researcher investment in science. We do not find strong evidence that past per researcher investments impact current output per researcher.

In most regressions the RE estimates of the elasticities are larger than the FE estimates. Not surprisingly, a Hausman test usually rejects the appropriateness of the random effects model. Thus, there is some evidence that the FE estimates are preferable to the RE estimates.

Sensitivity Analysis

We now investigate the robustness of the preceding results by examining a range of alternative specifications of our models. Recall that under every model specification in the previous section, the measures of scientific output were in natural logarithmic form. Our first task in assessing the robustness of our results is to examine their sensitivity to this non-linearity we imposed on our models. Panel A of table 7 replicates some of the main results of tables 4 and

5 except that scientific output is measured in levels rather than logs.[5] Generally, the estimates under this new specification are consistent with our original estimates.

Columns 1, 4, 7, and 10 of panel A in table 7 present the estimates of the unit change in scientific output with respect to a one percent increase in investment in science, omitting the time-varying country-specific covariates and the lags. For both measures of scientific output the FE and RE estimates are positive, economically significant, and precise. These results are consistent with our original estimates.

Columns 2, 5, 8, and 11 of panel A in table 7 present the estimates of the unit change in scientific output with respect to a one percent increase in investment in science, with the time-varying country-specific covariates included in the model. When the dependent variable is the number of researchers (columns 8 and 11), the results are again consistent with our original estimates. The RE and FE estimates are positive, economically significant, and precisely estimated. In contrast, when the dependent variable is the number of articles (columns 2 and 5), the results are seemingly puzzling. Indeed, both the RE and FE estimates are *negative* and economically significant. Moreover, the RE estimate is precisely estimated. Taken literally, these estimates indicate that an increase in total R&D expenditure is associated with a *decrease* in scientific output. This conclusion is neither plausible nor is it consistent with the rest of our estimates.

This apparent inconsistency is reconciled by adding a quadratic term of the natural logarithm of total R&D expenditure. Columns 3, 6, 9, and 12 of panel A in table 7 present the estimates of the change in scientific output with respect to a one percent increase in investment in science, with a quadratic investment term and the time-varying country-specific covariates included in the model. When the quadratic term is added, the RE and FE estimates are negative

[5] For brevity, we do not report the results for time dummies or covariates.

for the linear term and are positive for the quadratic term. This holds for both measures of scientific output. Importantly, for both measures of scientific output, the RE and FE estimates are much more precise when a quadratic term is added to the model. This suggests that there is some minimum amount of total R&D expenditure (convex expenditure) that must be attained before the relationship between scientific output in levels is positively related to the natural logarithm of total R&D expenditure. Therefore, these results are generally consistent with our original results. There is a positive (at least after some level of R&D expenditure) statistically and economically significant relationship between scientific output and investment in science.

We now examine the sensitivity of our results to our choice of measurement for investment in science by considering several alternative measures. First, we consider R&D expenditure as a percentage of GDP. Panel B of table 7 replicates some of the main results of tables 4 and 5, except that the natural logarithm of total R&D expenditure is replaced by R&D expenditure as a percentage of GDP. The results are generally consistent with our original estimates. Indeed, for every model specification, the RE and FE estimates of the percent change in scientific output associated with a one point increase of R&D expenditure as a percent of GDP are positive and economically and statistically significant. Moreover, the RE estimates tend to be larger than the FE estimates, which is also consistent with our original results.

Second, we consider the disaggregated components of R&D as measures of investment in science. The total R&D expenditure data from the OECD can be broken down into four components: 1) government R&D expenditure, 2) business R&D expenditure, 3) higher education R&D expenditure, and 4) private non-profit R&D expenditure. Table 8 replicates some of the main results in tables 4 and 5, except that it uses the various disaggregated components of total R&D expenditures as the measure of investment in science. Once again, these results are generally consistent with our original estimates. For both measures of scientific output, the RE and FE estimates are both positive and economically and statistically significant or they are not statistically significant. Moreover, the RE estimates tend to be larger than the FE estimates, which is consistent with previous results.

Analytical Problems

Though we believe the preceding results to be plausible, it is imperative that we discuss the major shortcomings in our models and in our data. First, the data set is an unbalanced panel. If attrition from the data set is correlated with the idiosyncratic error terms in our models, then we have a sample selection problem, which can cause biased estimators. However, when we use the FE estimator attrition *can* be correlated with country fixed effects. This is another reason to prefer our FE estimates over our RE estimates. Our model may also suffer from various forms of endogeneity such as omitted variables and simultaneity. This, of course, will also cause our estimators to be biased.

Conceptually, the largest problem with our analysis is that the quality of the scientific outputs is not taken into account. For instance, though more scientific investment is associated with more articles being produced, it is not clear whether additional articles have any real value. Traditionally, the best way to control for the quality of scientific output is to use the number of citations per article or the number of citations per researcher. Unfortunately, there are no systematic citation data available at the country-level. Thus, even if we could make the claim that increased investment does cause output to increase, we do not know whether the new output is a source of useful new ideas and thus able to contribute to economic growth.

Another conceptual problem arises when we consider our measure of output per researcher: articles per researcher. Presumably, increasing R&D expenditure per researcher will have two main effects. First, existing researchers' productivity will increase. Second, more people will decide to become researchers. If we make the plausible assumption that these new researchers are, on average, less productive than existing researchers, *measured* scientific output per researcher will *decrease* as the new, less productive, researchers become employed. Thus, the first effect increases measured output per researcher while the second effect decreases measured output per researcher. Whether aggregate output per researcher increases or decreases depends on which effect dominates. Our positive estimates of scientific output per researcher indicate that the first effect dominates. Moreover, since our model relating scientific output per researcher to per researcher investment in science does not take the second effect into account, the estimates in table 5 can be considered lower bounds of the elasticity of scientific output per researcher with respect to per researcher investment in science. It is also worth noting that this conceptual problem could account for the lack of evidence that past per researcher investment in science impacts current scientific output per researcher. Recall that the FE estimate of the elasticity of articles per researcher with respect the five-year averaged lag in column 8 of table 6 is not statistically different from zero. This could have resulted from the two effects roughly cancelling each other out.

Causality

Finally, it is imperative to emphasize that our task in this paper is description of the relationship between investment in science and scientific output. We are not making causal claims. Although we are able to control for all country fixed effects, there may be omitted time-varying covariates that complicate causal interpretation of our results.

Though we make no attempt to empirically assess causality, it is worth considering some possible mechanisms by which our results could have emerged. The first and most obvious interpretation of our results is that if a country increases its investment in science, this increases the ability of universities, businesses, and governments to hire more researchers, thus increasing the number of researchers in that country. Moreover, increased investment in science could enable current researchers to produce more articles and new researchers to produce articles that would not have otherwise existed. Under this mechanism, more investment causes more output and increased productivity.

Though the above interpretation is attractive, it is also simple to devise a story of reverse causality in which increases in the scientific output or scientific productivity cause increases in scientific investment. If a country has a large stock of researchers with the ability to productively exploit additional funding to produce useful research, that country has a strong incentive to invest in science. In addition, if the researchers in a country are highly productive, and produce many useful articles, the incentive to invest is even stronger. Under this mechanism, more output and higher productivity cause increased investment.

Finally, a cursory glance at the countries that score well in terms of large scientific outputs and large investments in science reveals that these are often the same countries that score well on a variety of social and economic indicators. We are able to control for country fixed effects and some time-varying country-specific effects, but this does not preclude the possibility that some other time-varying country-specific effect is causing both higher outputs and more investment in science. Under this mechanism more output, and higher investment in science are caused by some other variable. In sum, given the observational nature of our data and the lack of an attractive instrumental variable, it is impossible to know for sure in what direction causality runs. This is why the task of our paper is the more modest one of description.

VII. Conclusions

Economists have long emphasized that the generation of new ideas is essential for promoting economic growth. Clearly, formal scientific discovery is one important mechanism by which new ideas are generated. Therefore, it is extremely important for economists to understand the fundamental question of what promotes scientific discovery. This is all the more important given the fact that policymakers and members of the business community have been devoting larger shares of budgets to investments in science.

Toward the goal of understanding what promotes scientific discovery, this paper has attempted to analyze, in a cross-national context, the association between investments in science and scientific output. Across a variety of model specifications and estimation techniques, we consistently find a large, positive, statistically significant relationship between investment and output. A sensitivity analysis revealed that these results are robust against many departures from our original models. As noted our estimates of the elasticity of scientific output with respect to total investment in science range between 0.25 and 0.71. In addition, there is evidence past investments are positively associated with current output, even holding current investments constant. Our estimates of the elasticity of scientific output per researcher with respect to per researcher investment in science range between 0.30 and 0.43. In contrast to scientific output, we find little evidence that past investments impact current output per researcher.

Though these results are certainly plausible, a range of problems plague our analysis. These include the unbalanced panel structure of our data, endogeneity, and a lack of control for the quality of scientific outputs. These are major problems and the reliability of our estimates must be discounted accordingly. Moreover, the direction of causality is not readily apparent.

Clearly, much work remains to be done. Finding a satisfactory instrumental variable would help solve the endogeneity problem. In addition, better data could reduce the bias in the estimators. Nevertheless, the estimates presented in this paper are an important first step in understanding the relationship between investment in science and scientific output in a cross-national context.

References

Adams, J.D., & Griliches, Z. (1996). Research Productivity in a System of Universities. *NBER Working Paper No. 5833*.

Boardman, C. (forthcoming).

Coupé, T. (2003). Science Is Golden: Academic R&D and University Patents. *The Journal of Technology Transfer, 28*(1), 31-46.

Crespi and Geuna (2005). Modeling and Measuring Scientific Production: Results for a Panel of OECD Countries. *SPRU Electronic Working Paper Series*, Paper No. 133.

Crespi and Geuna (2004). The Productivity of Science. Report prepared for the Office of Science and Technology, Department of Trade and Industry.

Diamond, A.M. (1996). The Economics of Science. *Knowledge and Policy, 9*, 6-49.

Fox, M.F. (1983). Productivity among Scientists: A Critical Review. *Social Studies of Science, 12*(2), 285-305.

Hausman, J.A. (1978). Specification Tests in Econometrics. *Econometrica, 46*, 1251-1271.

Griliches, Z. (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics, 10,* 92-116.

Long, J.S., & McGinnis, R. (1981). Organizational Context and Scientific Productivity. *American Sociological Review, 46*(4), 422-442.

Lucas, R.E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics, 22,* 3-42.

Merton, R.K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Payne, A.A., & Siow, A. (1998). Estimating the Effects of Federal Funding on Universities using Alumni Representation on Congressional Appropriations Committees.

Pelz, D.C., Andrews, F.M. (1966/1976). *Scientists in Organizations* (Revised ed.). Ann Arbor: Institute for Social Research The University of Michigan.

Ramsden, P. (1994). Describing and Explaining Research Productivity. *Higher Education, 28*(2), 207-226.

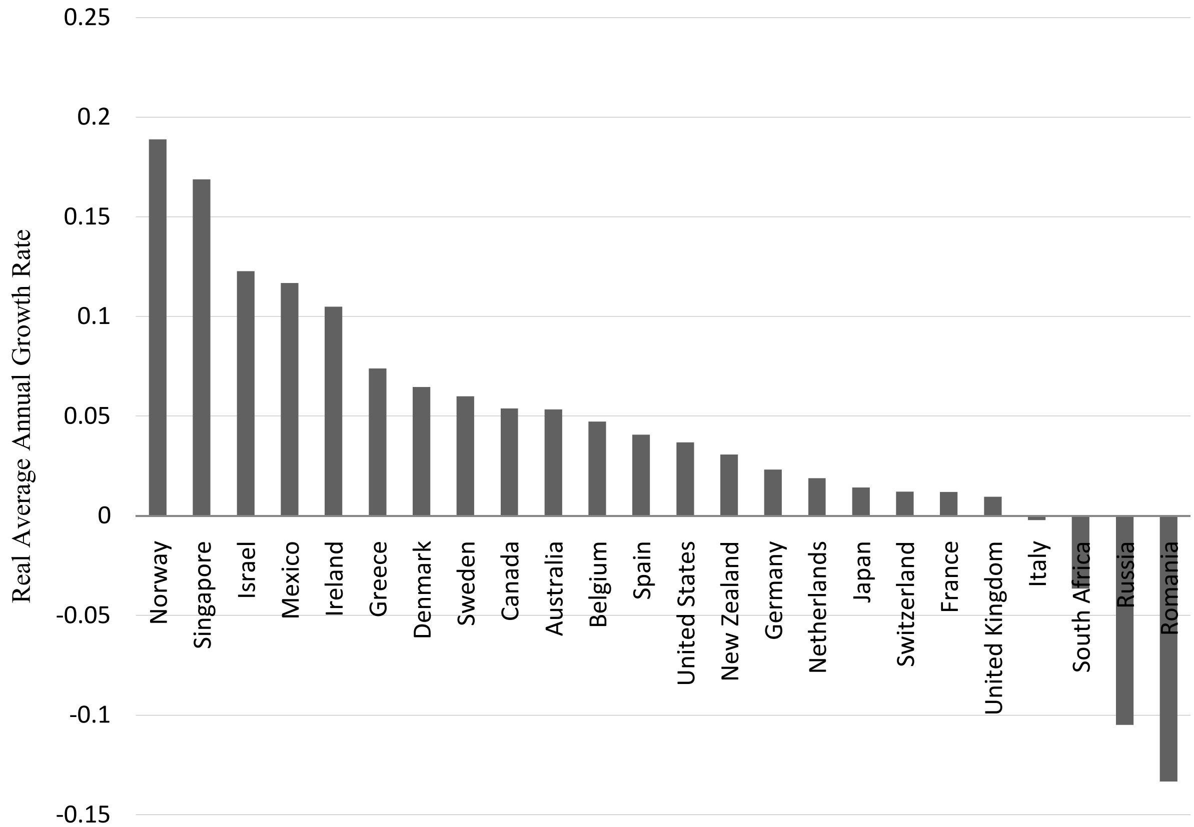
Romer, P. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy, 94*(5), 1002-1037.

Stephan, P.E. (1996). The Economics of Science. *Journal of Economic Literature, 34*(3), 1199-1235.

Teodorescu, D. (2000). Correlates of Faculty Publication Productivity: A Cross-National Analysis. *Higher Education*, 391(2), 201-222.

Weinberg, B. (2008). Scientific Productivity. Working Paper.

Figure 1: Real Average Annual Growth Rates in Total R&D Expenditure between 1990 and 2000



*Note*s: Author's calculations; data from the OECD's Main Science and Technology Indicators

Table 1: List of countries included in regressions

1. Algeria	29. Greenland	57. Norway
2. Argentina	30. Guatemala	58. Pakistan
3. Australia	31. Hong Kong, China	59. Panama
4. Austria	32. Hungary	60. Paraguay
5. Belgium	33. Iceland	61. Peru
6. Bolivia	34. India	62. Poland
7. Brazil	35. Indonesia	63. Portugal
8. Brunei Darussalam	36. Ireland	64. Romania
9. Bulgaria	37. Israel	65. Russian Federation
10. Burkina Faso	38. Italy	66. Seychelles
11. Cambodia	39. Japan	67. Singapore
12. Canada	40. Kazakhstan	68. Slovak Republic
13. Chile	41. Korea, Rep.	69. Slovenia
14. China	42. Kuwait	70. South Africa
15. Columbia	43. Latvia	71. Spain
16. Costa Rica	44. Lesotho	72. Sri Lanka
17. Croatia	45. Lithuania	73. Sweden
18. Cyprus	46. Luxembourg	74. Switzerland
19. Czech Republic	47. Macao, China	75. Thailand
20. Denmark	48. Macedonia, FYR	76. Tunisia
21. Ecuador	49. Madagascar	77. Turkey
22. El Salvador	50. Malaysia	78. United Kingdom
23. Estonia	51. Malta	79. United States
24. Ethiopia	52. Mexico	80. Uruguay
25. Finland	53. Myanmar	81. Venezuela, RB
26. France	54. Netherlands	82. Vietnam
27. Germany	55. New Zealand	83. Zambia
28. Greece	56. Nicaragua	

Notes: Due to the unbalanced structure of the data, the number of countries and the number of observations used to compute the estimates vary depending on the variables included in the model. No single model includes every country listed. See tables 4, 5, and 6 for details on the number of countries included in each regression.

Table 2: Description of variables

	Definition	Source
Scientific output		
Articles	Scientific and technical journal articles refer to the number of scientific and engineering articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedial research, engineering and technology, and earth and space sciences.	World Bank, World Development Indicators data set
Number of R&D researchers	Researchers in R&D are people trained to work in any field of science who are engaged in professional R&D activity. Most such jobs require completion of tertiary education.	World Bank, World Development Indicators data set
Articles per researcher	Articles per researcher are the number of scientific and technical journal articles divided by the number of researchers in R&D.	Author calculations
Investment in science		
Total R&D expenditure	Research and development is a term covering three activities: basic research, applied research, and experimental development. Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view. Applied research is original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective. Experimental development is systematic work, drawing on existing knowledge gained from research and/or practical experience, that is directed to producing new materials, products or devices; to installing new processes, systems and services; or to improving substantially those already produced or installed.	OECD, Main Science and Technology Indicators data set
R&D expenditure per researcher	R&D expenditure per researcher is total R&D expenditure divided by the number of researchers in R&D.	Author calculations
Covariates		
Public expenditure on education (% of GDP)	Public expenditure on education consists of public spending on public education plus subsidies to private education at the primary, secondary, and tertiary levels.	World Bank, World Development Indicators data set
Population	Population is the total number of people living within a certain geographic region	World Bank, World Development Indicators data set
GDP per capita	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current international dollars.	World Bank, World Development Indicators data set
Military expenditure (millions)	Military expenditures are based on the NATO definition, which includes all current and capital expenditures on the armed forces, including peacekeeping forces; defense ministries and other government agencies engaged in defense projects; paramilitary forces, if these are judged to be trained and equipped for military operations; and military space activities. Such expenditures include military and civil personnel, including retirement pensions of military personnel and social services for personnel; operation and maintenance; procurement; military research and development; and military aid (in the military expenditures of the donor country). Excluded are civil defense and current expenditures for previous military activities, such as for veterans' benefits, demobilization, conversion, and destruction of weapons. This definition cannot be applied for all countries, however, since that would require much more detailed information than is available about what is included in military budgets and off-budget military expenditure items. (For example, military budgets might or might not cover civil defense, reserves and auxiliary forces, police and paramilitary forces, dual-purpose forces such as military and civilian police, military grants in kind, pensions for military personnel, and social security contributions paid by one part of government to another.)	World Bank, World Development Indicators data set

Table 3: Summary statistics

Scientific output	Obs.	Mean	Standard Deviation	Minimum	Maximum
Articles	1,991	4,906	19,357	0	205,320
Number of R&D researchers	626	78,792	206,000	5.981	1,397,095
Articles per researcher	478	0.146	0.150	0	1.584
Investment in science					
Total R&D expenditure	592	19,206	46,214	61.27	325,305
R&D expenditure per researcher	332	0.128	0.060	0.015	0.271
Covariates					
Public expenditure on education (% of GDP)	1,360	4.698	2.117	1	18
Population (millions)	3,856	29.253	115.6	0.015	1,325
GDP per capita	3,440	6662	10,292	62	72,637
Military expenditure (millions)	2,677	605,467	4,215,800	0.00057	86,500,000

Table 4: Relationship between log total R&D expenditure and log total articles

	Dependent variable: Log(total articles)							
	Random effects				Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(total R&D expenditure)	0.707 (0.038)***	0.453 (0.070)***	0.432 (0.132)***	0.227 (0.089)**	0.246 (0.058)***	0.419 (0.091)***	0.354 (0.130)***	0.162 (0.104)
Log(total R&D expenditure) (t-1)			0.031 (0.193)				-0.008 (0.183)	
Log(total R&D expenditure) (t-2)			-0.185 (0.187)				-0.140 (0.177)	
Log(total R&D expenditure) (t-3)			0.093 (0.172)				-0.017 (0.170)	
Log(total R&D expenditure) (t-4)			0.024 (0.149)				0.102 (0.145)	
Log(total R&D expenditure) (t-5)			0.270 (0.093)***				0.309 (0.098)***	
Log(total R&D expenditure) (5 year-averaged lag)				0.401 (0.068)***				0.344 (0.090)***
Log(public expenditure on education as % of GDP)		0.420 (0.078)***	0.233 (0.074)***	0.197 (0.075)***		0.440 (0.083)***	0.195 (0.078)**	0.192 (0.082)**
Log(population)		0.365 (0.079)***	0.157 (0.063)**	0.188 (0.073)**		-0.274 (0.407)	-0.891 (0.579)	-0.078 (0.567)
Log(GDP per capita)		0.345 (0.094)***	0.155 (0.078)**	0.185 (0.089)**		0.057 (0.196)	-0.159 (0.175)	-0.078 (0.181)
Log(military expenditure) [centered]		0.066 (0.017)***	0.082 (0.024)***	0.101 (0.025)***		0.055 (0.027)*	0.138 (0.033)***	0.118 (0.033)***
Log(military expenditure)^2 [centered]		-0.015 (0.002)***	-0.016 *0.003)***	-0.019 (0.003)***		-0.014 (0.004)***	-0.009 (0.006)	-0.022 (0.005)***
Constant	2.482 (0.318)***	-5.130 (1.640)***	-1.255 (1.252)	-1.686 (1.449)	6.234 (0.468)***	8.546 (7.247)	20.154 (10.074)**	6.391 (9.909)
R^2 within	0.364	0.690	0.658	0.623	0.422	0.701	0.681	0.637
R^2 between	0.869	0.941	0.973	0.971	0.824	0.444	0.020	0.946
R^2 Overall	0.880	0.941	0.969	0.968	0.758	0.537	0.074	0.942
N	496	268	199	199	496	268	199	199
Countries	37	35	27	27	37	35	27	27
Hausman (Prob>chi2)	0.000	0.101	0.003	0.524				

Notes: Standard errors in parentheses below each coefficient. (*) significant at 10%; (**) significant at 5%; (***) significant at 1%. All models include year effects. Due to the unbalanced structure of the data, the number of countries and number of observations used to compute the estimates vary depending on the variables included in the model.

Table 5: Relationship between log total R&D expenditure and log total number of R&D researchers

	Dependent variable: Log(number of R&D researchers)							
	Random effects				Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(total R&D expenditure)	0.629 (0.036)***	0.679 (0.080)***	0.649 (0.138)***	0.536 (0.092)***	0.513 (0.043)***	0.511 (0.086)***	0.558 (0.129)***	0.432 (0.093)***
Log(total R&D expenditure) (t-1)			0.074 (0.196)				-0.005 (0.175)	
Log(total R&D expenditure) (t-2)			-0.264 (0.174)				-0.206 (0.155)	
Log(total R&D expenditure) (t-3)			0.284 (0.176)				0.257 (0.156)	
Log(total R&D expenditure) (t-4)			-0.124 (0.156)				-0.028 (0.139)	
Log(total R&D expenditure) (t-5)			0.262 (0.095)***				0.115 (0.094)	
Log(total R&D expenditure) (5 year-averaged lag)				0.314 (0.071)***				0.220 (0.079)***
Log(public expenditure on education as % of GDP)		0.133 (0.068)**	0.083 (0.071)	0.059 (0.072)		0.095 (0.066)	0.074 (0.069)	0.059 (0.069)
Log(population)		0.322 (0.092)***	0.109 (0.086)	0.140 (0.085)		2.721 (0.452)***	2.181 (0.526)***	2.398 (0.496)***
Log(GDP per capita)		-0.218 (0.106)**	-0.446 (0.098)***	-0.420 (0.098)***		0.081 (0.171)	0.154 (0.191)	0.155 (0.190)
Log(military expenditure) [centered]		-0.076 (0.033)**	-0.035 (0.033)	-0.036 (0.033)		-0.108 (0.038)***	-0.082 (0.039)**	-0.084 (0.039)**
Log(military expenditure)^2 [centered]		0.007 (0.005)	0.004 (0.004)	0.004 (0.004)		0.003 (0.006)	-0.002 (0.007)	-0.002 (0.007)
Constant	5.299 (0.339)***	1.258 (1.884)	5.567 (1.678)***	5.093 (1.660)***	7.636 (0.363)***	-40.695 (8.132)***	-33.91 (9.147)***	-37.341 (8.693)***
R^2 within	0.722	0.685	0.722	0.699	0.728	0.739	0.763	0.752
R^2 between	0.859	0.912	0.952	0.950	0.858	0.773	0.854	0.842
R^2 Overall	0.844	0.916	0.946	0.945	0.840	0.787	0.862	0.851
N	332	221	183	183	332	221	183	183
Countries	36	34	26	26	36	34	26	26
Hausman (Prob>chi2)	N/A	0.000	0.001	0.000				

Notes: Standard errors in parentheses below each coefficient. (*) significant at 10%; (**) significant at 5%; (***) significant at 1%. All models include year effects. Due to the unbalanced structure of the data, the number of countries and number of observations used to compute the estimates vary depending on the variables included in the model. The Hausman test comparing columns 1 and 5 was not applicable because the model fitted on these data did not meet the asymptotic assumptions of the test.

Table 6: Relationship between log per researcher R&D expenditure and log articles per researcher

	Dependent variable: Log(articles per researcher)							
	Random effects				Fixed Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log(total R&D expenditure per researcher)	0.476 (0.059)***	0.434 (0.071)***	0.420 (0.121)***	0.352 (0.118)***	0.430 (0.066)***	0.304 (0.088)***	0.375 (0.130)***	0.303 (0.112)**
Log(total R&D expenditure per researcher) (t-1)			-0.184 (0.139)				-0.245 (0.114)*	
Log(total R&D expenditure per researcher) (t-2)			-0.042 (0.167)				-0.046 (0.176)	
Log(total R&D expenditure per researcher) (t-3)			0.186 (0.184)				0.174 (0.192)	
Log(total R&D expenditure per researcher) (t-4)			-0.046 (0.152)				-0.083 (0.156)	
Log(total R&D expenditure per researcher) (t-5)			0.326 (0.116)***				0.181 (0.127)	
Log(total R&D expenditure per researcher) (5 year-averaged lag)				0.295 (0.140)**				-0.005 (0.183)
Log(public expenditure on education as % of GDP)		0.119 (0.077)	0.085 (0.106)	0.051 (0.016)		0.052 (0.082)	0.096 (0.117)	0.066 (0.113)
Log(population)		-0.133 (0.045)***	0.016 (0.071)	0.007 (0.064)		-1.110 (0.596)*	-0.256 (1.017)	-0.097 (1.019)
Log(GDP per capita)		0.079 (0.066)	-0.125 (0.116)	-0.098 (0.111)		-0.658 (0.191)***	-0.707 (0.368)*	-0.792 (0.359)**
Log(military expenditure) [centered]		0.057 (0.033)	-0.071 (0.056)	-0.055 (0.053)		0.173 (0.047)***	0.008 (0.139)	0.053 (0.138)
Log(military expenditure)^2 [centered]		-0.018 (0.005)***	-0.004 (0.007)	-0.007 (0.007)		-0.005 (0.008)	0.009 (0.002)	0.009 (0.012)
Constant	-1.196 (0.273)***	0.426 (1.158)	0.413 (1.695)	0.332 (1.571)	-1.367 (0.284)***	23.470 (10.511)**	9.566 (18.06)	7.485 (18.246)
R^2 within	0.297	0.259	0.215	0.113	0.298	0.346	0.276	0.205
R^2 between	0.339	0.788	0.796	0.790	0.338	0.069	0.113	0.324
R^2 Overall	0.459	0.782	0.799	0.793	0.458	0.009	0.117	0.342
N	308	207	107	107	308	207	107	107
Countries	36	34	24	24	36	34	24	24
Hausman (Prob>chi2)	0.001	0.005	0.555	0.109				

Notes: Standard errors in parentheses below each coefficient. (*) significant at 10%; (**) significant at 5%; (***) significant at 1%. All models include year effects. Due to the unbalanced structure of the data, the number of countries and number of observations used to compute the estimates vary depending on the variables included in the model.

Table 7.A: Relationship between log total R&D expenditure and two measures of scientific output in levels

	Random effects			Fixed Effects			Random Effects			Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
	Articles						Number of R&D researchers (thousands)					
Log(total R&D expenditure)	2224.9 (535.5)***	-2125 (1058)**	-26221 (3653)***	1834.2 (5241)***	-1614 (1119)	-13547 (3549)***	140.19 (12.53)***	32.83 (12.55)***	-344.6 (40.8)***	158.01 (16.4)***	24.31 (12.45)**	-256.24 (45.26)***
Log(total R&D expenditure)^2			1649.8 (238.8)***			832.9 (235.8)***			23.9 (2.48)***			17.99 (2.81)***
R^2 within	0.288	0.370	0.359	0.289	0.378	0.412	0.329	0.391	0.511	0.331	0.409	0.523
N	497	268	268	497	268	268	332	221	221	332	221	221
Countries	38	35		38	35	35	36	34	34	36	34	34
Covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Table 7.B: Relationship between total R&D expenditure as a percentage of GDP and two measures of scientific output

Panel B												
	Log(articles)						Log(number of R&D researchers)					
R&D expenditure as % of GDP	0.182 (0.027)***	0.152 (0.029)***		0.153 (0.026)***	0.064 (0.027)**		0.268 (0.034)***	0.183 (0.038)***		0.238 (0.033)***	0.107 (0.033)***	
R^2 within	0.399	0.355		0.400	0.478		0.388	0.491		0.389	0.620	
N	606	378		606	378		558	335		558	335	
Countries	80	68		80	68		82	64		82	64	
Covariates	No	Yes		No	Yes		No	Yes		No	Yes	

Notes: Standard errors in parentheses below each coefficient. (*) significant at 10%; (**) significant at 5%; (***) significant at 1%. All models include year effects. Due to the unbalanced structure of the data, the countries and number of observations vary depending on the variables included in the model.

Table 8: Relationship between log of disaggregated measures of total R&D expenditure and two measures of scientific output.

	Random Effects		Fixed Effects	
	(1)	(2)	(3)	(4)
Log(articles) as dependent variable				
Log(government)	0.581	0.104	0.321	0.102
	(0.043)***	(0.053)*	(0.052)***	(0.059)*
	R^2=0.416	R^2=0.644	R^2=0.439	R^2=0.678
	n=498	n=271	n=498	n=271
Log(business)	0.491	0.213	0.167	0.179
	(0.033)***	(0.046)***	(0.039)***	(0.057)***
	R^2=0.354	R^2=0.669	R^2=0.417	R^2=0.685
	n=502	n=274	n=502	n=274
Log(higher education)	0.526	0.129	0.073	0.096
	(0.036)***	(0.063)**	(0.073)	(0.066)
	R^2=0.302	R^2=0.664	R^2=0.387	R^2=0.674
	n=494	n=271	n=494	n=271
Log(private Non profit)	0.067	0.005	0.028	0.006
	(0.025)***	(0.022)	(0.024)	(0.024)
	R^2=0.335	R^2=0.515	R^2=0.361	R^2=0.536
	n=335	n=194	n=335	n=194
Covariates	No	Yes	No	Yes
Log(number of R&D researchers) as dependent variable				
Log(government)	0.461	-0.012	0.109	-0.075
	(0.039)***	(0.056)	(0.041)***	(0.056)
	R^2=0.552	R^2=0.603	R^2=0.601	R^2=0.688
	n=332	n=221	n=332	n=221
Log(business)	0.431	0.352	0.368	0.249
	(0.027)***	(0.043)***	(0.028)***	(0.045)***
	R^2=0.744	R^2=0.684	R^2=0.747	R^2=0.732
	n=332	n=221	n=332	n=221
Log(higher education)	0.190	0.095	0.087	0.148
	(0.036)***	(0.064)	(0.037)**	(0.057)***
	R^2=0.589	R^2=0.613	R^2=0.599	R^2=0.696
	n=332	n=221	n=332	n=221
Log(private Non profit)	-0.003	-0.025	-0.019	-0.001
	(0.016)	(0.019)	(0.015)	(0.017)
	R^2=0.677	R^2=0.628	R^2=0.679	R^2=0.719
	n=229	n=159	n=229	n=159
Covariates	No	Yes	No	Yes

Notes: Standard errors in parentheses below each coefficient. (*) significant at 10%; (**) significant at 5%; (***) significant at 1%. All models include year effects. Due to the unbalanced structure of the data, the countries and number of observations vary depending on the variables included in the model. R^2 is the within coefficient of determination.